PRACTICE UNIT 7 – CORRELATION AND LINEAR REGRESSION MODEL

OBJECTIVES

The objectives of this practice are:

- Analyse data obtained in a survey and strengthen the use of the Excel spreadsheet.
- Analyse the degree of linear association (correlation) between two variables.
- Fit a linear model using the least squares method. Measure the goodness of fit of a linear model numerically and graphically. Use an estimated linear model to predict.
- Learn and use the Excel functions that allow the correlation and regression analysis between two variables.

EXERCISE U7_1. (Sheet '*Height-Weight*')

From the survey conducted with the students of the course Statistics I (same of previous practices) we have selected the values of their *Height* and *Weight*.

a) Use a scatter plot to present graphically the variables *Height* and *Weight*. What type of association does the plot suggest?

b) Calculate and interpret the covariance. Calculate a measure that indicates the intensity and the type of the linear association between the two variables.

c) Using Excel functions, obtain the least squares regression line that explains the *weight* given the *height*. Interpret the parameters of the estimated regression line. Show this regression line in the graphic presentation obtained in a).

d) Analyse the goodness of fit of the previous regression model numerically. Interpret the measure employed.

e) Calculate the residuals and analyse the goodness of fit graphically.

f) Analyse the goodness of fit using the general coefficient of determination.

g) Obtain the least squares regression line that explains the *height* given the *weight*.

h) Using the appropriate linear fit (the one estimated in c) or the one estimated in g)) answer to the following questions: What would be the weight of a student 170 cm height? And that of a student 205 cm height? Are these estimations reliable?

EXERCISE U7_2. (Sheet '*Advertising*')

The managers of a multinational company want to analyse the possible relationship between the Annual profits (Y) and the Advertising expenditure (X) for a group of various products whose values are shown in the table.

Year	X (million €)	Y (million €)
1	2	-6
2	2.8	-3
3	3.9	0
4	4.2	3
5	5.8	6
6	6.2	9
7	7.5	12
8	8.2	15
9	9.3	20
10	10.9	25

a) Calculate some numerical measures to quantify the intensity of the linear association between the two variables. Interpret the values.

b) Determine the best fit between the annual profits and the advertising expenditure using a linear function. Interpret the parameters of the estimated regression line.

c) Analyse the goodness of fit of the linear regression model numerically and graphically.

d) Estimate the annual profits of the company if it spends 15 million euros in advertising. Is this prediction reliable?